Data Compression Method and Apparatus

Abstract of Disclosure

An improved data compression method and apparatus is disclosed, particularly for compressing large database tables. A data structure is disclosed which is fully compatible with the traditional DBMS demands, including the random access requirement of RDBMS. The data structure is built on a mixed format physical layout comprising of fixed-sized fields and variable-sized fields which are compressed depending on the size and frequency of the fields. An improved compression ratio is achieved by exploiting redundancy in the mixed format physical layout to encode the column-wise redundancy in the data itself and the correlations among columns. The present invention provides a very fast random access decompression and enables not only greater compression ratios, but also permits flexibility of choosing from a number of compression algorithms.

Figures